Speech production and phonetics

Slides for this lecture are partly based on those created by Katariina Mahkonen for TUT course ”Puheenkäsittelyn menetelmät” in Spring 2013.
Books: Speech Communications, Douglas O'Shaughnessy

Speech production anatomy
» Overview, source-filter model of speech production
» Vocal tract
» Larynx, glottis
Articulatory phonetics
» Vowels
» Consonants
» International phonetic alphabet
What is phonetics?

Phonetics studies speech:
- Production -> ARTICULATORY
- Acoustic realization -> ACOUSTIC
- Perception -> AUDITORY
Vocal organs

» Vocal organs can be subdivided into:

- **central** (Broca’s area, Wernicke’s area)
and peripheral
Source-filter model of speech production

» Speech production can be viewed as acoustic filtering operation

» Larynx (vocal folds) and lungs provide source excitation

» Vocal tract acts as a filter that shapes the spectrum of the speech signal
Vocal tract

- Vocal tract refers to vocal organs after the larynx
- Divided into following sections:
  - Pharynx cavity
  - Nasal cavity
  - Oral cavity
- Organs of vocal tract that move to produce various speech sounds
  - Tongue
  - Soft palate (velum) -> opens/closes path to nasal cavity
  - Lower jaw
  - Lips
Vocal tract and Formants

- Vocal tract acts like an adjustable filter: resonant frequencies are determined by the vocal tract shape.
1. Nasal cavity
2. Hard palate
3. Alveolar ridge
4. Soft palate (velum)
5. Tongue tip
6. Dorsum
7. Uvula
8. Radix
9. Pharynx
10. Epiglottis
11. False vocal cords
12. Vocal cord (vocal fold)
13. Larynx
14. Esophagus (=gullet) → to stomach
15. Trachea (=windpipe) → to lungs

closes off larynx while eating
opens nose cavity for m, n, ng
MRI (Magnetic Resonance Imaging) images of the vocal tract

http://personal.ee.surrey.ac.uk/Personal/P.Jackson/Nephthys/jaleel.html
Glottis (in larynx)

- Glottis is the space between vocal folds
- From the speech production viewpoint, the role of larynx is to turn the silent flow of air from the lungs into audible sound
- The arytenoid cartilages are a pair of small three-sided pyramids which form part of the larynx, to which the vocal folds (vocal cords) are attached
  - Muscle that controls the vocal folds
    - Tightness
    - Position

http://www.youtube.com/watch?v=wjRsa77u6OU
Function of the vocal folds

A: vocal folds and arytenoids closed -> glottal closure (no airflow)
B: Vocal folds vibrating, arytenoids closed -> phonation, f0; voicing
C: Vocal folds close, arytenoids open -> whisper
D: glottal constriction -> weak unvoiced noise, glottal fricative [h]
E: rest/breathing position -> unvoiced consonants
F: deep-breath position (sigh / breathlessness) -> not used for speech
Sources of sound energy

» Vocal fold vibration
  > Is caused by pressurized air passing through the membranous portion of the narrowed glottis.
  > Causes repeated opening and closing of the glottis
  > Formation of voiced sounds in this way is called phonation
  > Frequency of vibration: fundamental frequency $F_0$ can be altered with muscles from 80-400 Hz for males, 120-800 Hz for females, 300 Hz for children.

» Turbulence
  > Air moving quickly through a small hole
  > Fricative or unvoiced sounds
  > E.g. tongue/teeth (“ss” in “hiss”)

» Explosion
  > Release of pressure build up
  > E.g. behind lips (“p” in “peak”) or tongue (“t” in “tell”)
  > Plosive sounds

Compare “b” in “bat” (voiced plosive) with “p” in “pat” (unvoiced plosive)
Articulatory phonetics and International Phonetic Alphabet
Articulatory phonetics

» One goal of phonetics is to classify phonemes of different languages

> Phonetic alphabets:
+ International phonetic alphabet (IPA)  (chart)
+ Represents sounds with symbols: For notational reasons (ASCII-based) others are used too, e.g. Arpabet

» Phonetics describes phonemes as accurately as possible based on their articulation
Classification of speech sounds

» Consonant vs. vowel:
  consonants involve an obstruction in air stream above the glottis.

» Voiced vs. voiceless:
  voiced if vocal chords vibrate

» Nasal vs. oral:
  nasal if air travels through nasal cavity and oral cavity closed

» Lateral vs. non-lateral:
  In lateral phonemes, air stream passes through the sides of the oral cavity ("ball", "lateral") and not through the middle
Vowels are voiced phonemes, where the vocal tract is open. Vowels are characterized by using articulation features:

- **Open-Close dimension** refers to how close the tongue is to the roof of the mouth. The more closer to palate the more ”closed” the the vowel is.
- **Front-Back dimension** refers to position of articulation by means of tongue positions: the narrowest point of the vocal tract is essential.
- **Lip roundedness** (binary value), right&left of bullet: rounded&unrounded.
- **Nasalization**
  - When the velum is open, airflow gets to the nasal cavity and a nasal phoneme is produced. When the velum is closed, an oral phoneme is produced.

[www.internationalphoneticalphabet.org/]
[ipa-sounds/ipa-chart-with-sounds/sound](www.internationalphoneticalphabet.org/)

[IPA Chart](www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/sound)
Consonants

» In most consonants, the airflow is obstructed at some point

» Consonants are characterized by:

1. **Voicing** – voiced or unvoiced
2. **Place of articulation**
3. **Manner of articulation**
Voicing of consonants

» Voicing is determined by the vibration of the vocal folds

» A consonant can be voiced or unvoiced

» In English, voiced consonants include [v] (van), [z] (zip), [ʒ] (confusion), [b], [d], [g], [dʒ] (gin)

» Unvoiced consonants include: [f], [s], [p], [t], [k], [h], [s], [tʃ]
Place of articulation tells where is the primary constriction along the vocal track.

Consonant’s places of articulation:

- **bilabial** (1): made with the two lips (P,B,M)
- **labio-dental** (2): lower lip & upper front teeth (F,V)
- **dental** (4): tongue tip/blade & upper front teeth (TH,DH)
- **alveolar** (5): tongue tip/blade & alveolar ridge (T,D,N)
- **retroflex**: tongue tip & back of the alveolar ridge (R)
- **palato-alveolar**: tongue tip & back of the alveolar ridge (SH)
- **palatal** (6): front of the tongue & hard palate (Y,ZH)
- **velar** (7): back of the tongue & soft palate (K,G,NG)
- **uvular** (8): back of the tongue against or near the uvula.
- **pharyngeal**: (9) in the pharynx
- **glottal**: (10) in the glottis

(you do not have to remember the above latin words)
Consonants’ manners of articulation

» Main variation in the manner of articulation regards the question how freely the air stream flows when the consonant is produced

» Sonorants: continuous, non-turbulent airflow in the vocal tract

» Obstruent: airflow is partly or completely obstructed
Sonorants sounds where the air stream passes unobstructed through the vocal tract (includes vowels and consonants)

- **Semivowels** (aka glides): vowel-like sounds with greater constriction than corresponding vowels (/y/, /w/: ”yes”, ”well”).
- **Liquids** have spectra similar to vowels, but few decibels weaker.
- **Lateral** (”led”): obstruction of the air stream at a point along the center of the oral tract, with incomplete closure between one or both sides of the tongue and the roof of the mouth (/l/)
- **Retroflex** (”red”): tip of the tongue is curled back slightly (/r/)
- **Nasal**: soft palate down, airflow is through the nasal tract (/m/, /n/)
- **Approximants** are similar to fricatives, but articulators do no come close enough to generate turbulent airflow.
Obstruents are consonants where the airflow is partly of completely obstructed at some point

» **Plosive**: complete obstruction with sudden (explosive) release (/p/, /b/, /t/, /d/, /k/, /g/)

» **Fricative**: articulators close together, turbulent airflow produced. Aperiodic, with usually most of the energy at high frequencies (/f/, /v/, /th/, /dh/, /s/, /z/, /sh/, /zh/, /h/)
Flaps and Trills

» In **trills** the articulator vibrates rapidly with frequency of 20-25 Hz against the place of articulation. Only English trill is /r/ as in “roar”, where tongue touches the alveolar ridge for two to three vibrations.

» In **flaps** the articulation organs touch only once by a single contraction of the muscles involved.
**IPA – international phonetic alphabet**

**THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)**

<table>
<thead>
<tr>
<th>Consonants (Pulmonic)</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plosive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>p b</td>
<td>t d</td>
<td>t d</td>
<td>c j</td>
<td>k g q g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Voiced in case of only one consonant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m m̃</td>
<td>n</td>
<td>ñ</td>
<td>η η̃</td>
<td>η η̃ Ñ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>b</td>
<td>r</td>
<td>η</td>
<td>η̃</td>
<td>r R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap or Flap</td>
<td>u v̄</td>
<td>r</td>
<td>r̃</td>
<td>r̃</td>
<td>r R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>Ø ß φ β f v θ ð s z</td>
<td>f z</td>
<td>s z</td>
<td>ç j x y χ k</td>
<td>h ħ</td>
<td>h h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral fricative</td>
<td>l l̃</td>
<td>l l̃</td>
<td></td>
<td>l l̃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>u</td>
<td>j</td>
<td>j̃</td>
<td>j̃</td>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>l l̃</td>
<td>l l̃</td>
<td></td>
<td>l l̃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Voiceless consonants on the left of left/right pair
Voiced in case of only one consonant
Other phonetics terms

- **Phoneme**: the smallest linguistic unit which may bring about a change of meaning (kill vs. kiss). Phonemes are combined to form larger entities such as words. Noted in text with slashes e.g. /i/
- **Phone**: individual spoken realization of a phoneme
  - In principle all phones are different
  - Different speech sounds that are realizations of the same phoneme are known as **allophones**
  - Noted in text with brackets e.g. [i]
- **Coarticulation**: vocal organs move in a continuous manner and therefore (conceptually isolated) speech sound is influenced by, and becomes more like, a preceding or following speech sound.
- **Diphone**: the time-span between the middle-part of a phone until the middle part of the following phone. Includes phone transition.
- **Triphone**: a temporal unit that covers two diphones.
Prosody refers to longer-term properties of speech.

- **Rhythm**: varying the temporal length of syllables (or some other units)

- **Stress**: relative emphasis of syllables in a word or certain words in a sentence, manifested in higher/lower pitch or dynamics (loudness)

- **Intonation**: variation of pitch over a segment of multiple words (e.g. Sentence) that may
  + indicate the attitudes and emotions of the speaker
  + signal the difference between statement and question
  + focus attention on the important words
Acoustically, speech signal, as any sound, can be viewed as air pressure level variation.

Acoustic phonetics studies the acoustic characteristics of speech and their relationships to the speech production.

Longitudinal waves:  http://www.kettering.edu/physics/drussell/Demos/waves/wavemotion.html
The vocal tract can be treated as an acoustic tube with resonance frequencies called *formants*, $F_i$, where $i$ is the formant order, and $i=1$ is the lowest frequency.
Speech production and modeling

Quatieri: Discrete –Time Speech Signal Processing  Principles  and Practice